心理科学进展　2018, Vol. 26, No. 9, 1535–1544
Advances in Psychological Science　DOI: 10.3724/SP.J.1042.2018.01535

·研究构想(Conceptual Framework)·

汉语言语产生的语音加工单元——基于音位的研究*

屈青青　刘维琳　李兴珊
(中国科学院心理研究所行为科学重点实验室, 北京 100101) (中国科学院大学心理学系, 北京 100049)

摘　要　言语产生的语音加工单元具有跨语言的特异性。在印欧语言中, 音位是语音加工的重要功能单元。音位指具体语言中能够区别意义的最小语音单位, 如“big”包含三个音位/b/, /i/, /g/。目前, 在汉语言语产生中, 对音位的研究较少。本项目拟采用事件相关电位技术, 对汉语言语产生中的音位加工进行探讨, 试图考察: 在汉语言语产生中, 1)音位加工的心理现实性, 以及音位表征是否受第二语言、汉语拼音习得、拼音使用经验的影响? 2)音位的加工机制是怎样的? 具体而言, 音位加工的特异性、位置编码、组合方式、时间进程是怎样的? 对这些问题的回答, 将有助于深化对汉语言语产生的认识, 为建立汉语言语产生计算模型提供基础; 为比较印欧语言与汉语在机制上的异同提供基础; 为制定汉语语音教育教学方法提供心理学依据。

关键词　言语; 汉语; 语音加工单元; 音位; 事件相关电位
分类号　B842

1　研究背景

言语产生, 作为将思想转化为口头言语表达的认知加工过程, 包含以下四个主要加工阶段: 1)概念准备, 即说话者首先需要明确试图表达的概念意图; 2)词条选择, 即为所准备的概念从心理词典中提取适当的词汇; 3)音韵编码, 即构建词汇的语音信息; 4)发声执行阶段, 将计划的词汇用外显的声音表达出来(Caramazza, 1997; Dell, 1986, 1988; Levelt, Roelofs, & Meyer, 1999; Rapp & Goldrick, 2000)。在言语产生领域, 音韵编码作为言语产生的关键过程, 近年来受到了广泛的关注。音韵编码所涉及的语音加工单元是最热门的核心科学问题之一。基于英语或荷兰语等印欧语言的理论/模型(Caramazza, 1997; Dell, 1986, 1988; Levelt, 1989; Levelt et al., 1999)均假设在言语产生的音韵编码系统中存在着抽象的音位表征(phonemic representation), 音位是语音信息的提取、编码的重要功能单元。

不同语言(或方言)在音韵特征上存在着显著差异(Ladefoged, 2001; Ladefoged & Maddieson, 1996)。这种音韵特征的差异很可能导致不同语言间语音表征的不同。研究发现, 与印欧语言不同, 音位在日语言语产生中的作用甚微, 而音拍(mora)的作用显著(Kureta, Fushimi, & Tatsumi, 2006)。总之, 现有研究提示, 语音加工单元具有跨语言的特异性。

目前, 针对汉语言语产生中的语音表征单元的研究已获得一些进展, 然而, 这些研究多集中在对“音节在汉语言语产生中的作用”的探讨, 对音位的实证研究还非常少。其实, 中国古代学者就开始了对汉语音节的下一级、更小单位的分析。在东汉时期, 人们发明了注音方法“反切”, 以两个字来表示一个音节的两部分: 声母和韵母。在明朝, 人们把汉语音节划分为类似音位的单元: 声母、韵头、韵腹、韵尾。直到上世纪, 中国学者开始将西方的音位学理论运用于汉语研究, 以音位(元音、辅音)分析普通话。那么, 在汉语产生中, 音位加工是否具有心理现实性呢? 如果有, 音位的加工机制又是怎样的呢? 这些重要的科学问题还有待探讨。

收稿日期: 2017-11-03
* 国家自然科学基金(31771212)。
通信作者: 屈青青, E-mail: quqq@psych.ac.cn

2 国内外研究现状

2.1 音位(phoneme)[1]

音位是具体语言中有区别词的语音形式作用的最小语音单位(叶蜚声, 徐通锵, 2010), 是按语音的辨义作用归纳出来的音类, 不同语言的音位系统不能进行对应(黄伯荣, 廖序东, 2011)。当两个最小的语音单位可以区分不同的词汇时, 它们即构成了不同的音位。比如, 在汉语普通话中, “米”和“你”就是依靠两个辅音/m/、/n/区分的, 具有区别词的语音形式, 区别词的意义的作用, 所以/m/、/n/归为两个音位。在英语中, /b/、/p/属于两个不同的音位, 因为他们可以区分“big”、“pig”; 在“sin”、“sing”中, 尾音/n/、/ŋ/构成了两个不同的音位。音位一般用两条斜线标明。音位[2]包含元音音位和辅音音位, 从元音中归纳出来的称为元音音位, 如/a/、/o/、/e/; 从辅音中归纳出来的称为辅音音位, 如/b/、/p/、/m/。不同语言所使用的音位数量有所差异, 在汉语普通话中, 存在 22 个辅音音位和 10 个元音音位(叶蜚声, 徐通锵, 2010, 详见表 1)。在英语中, 音位数量随方言而变化, 最多可达到 24 个辅音音位, 20 个元音音位。传统语言学理论认为每种语言内均包含一组较少数量的音位, 这些音位按照各种方式进行组合, 从而表征每种语言的数千个词汇(Chomsky & Halle, 1968; Trubetzkoy, 1969)。

2.2 印欧语系言语产生中音位的心理现实性

正如上文所述, 语音加工单元, 即在长时记忆中存储和提取语音形式的功能单元(Baudouin de Courtenay, 1972), 是言语产生领域的核心科学问题。现有的实验证据和模型均支持音位是印欧语言音韵编码的重要功能单元。支持音位为语音加工单元的最初证据来自于语误分析。研究者发现大多数的语误涉及单个音位的插入, 删除, 替换或转换(e.g., **Y**ork **l**ibrary→ **l**ork **y**ibrary, **r**eading list → **l**eading list; Dell, 1986; Garrett, 1980; Meringer & Mayer, 1895; Shattuck-Hufnagel, 1979)。而涉及整个音节(e.g., **nap**kin → kin**nap**)或单个语音特征(e.g., **b**lue → **p**lue)的语音错误却相当少(Shattuck-Hufnagel, 1979, 1983)。语误分析为音位的心理现实性提供了直接证据。

支持音位的心理现实性的证据还来自于实验研究。研究者采用内隐启动范式(implicit priming paradigm), 让被试先学习几个单词对(刺激词−反应词), 在测试阶段, 只呈现刺激词, 要求被试又快又准确地说出对应的反应词。在同源条件下, 同一个 block 内所有的反应词具有某种相同的词形信息, 异源条件的反应词无关。研究发现, 当反应词之间的重叠部分出现在词首时, 同源组的反应时间快于异源组(Meyer, 1990; Roelofs, 1996, 1998; 张清芳, 2008)。Meyer (1991)利用该范式, 操控反应词之间是否存在首音位重叠。结果表明, 当反应词之间存在首音位重叠时(e.g., /d/: **d**ans-**d**op-**d**eugd-**d**oek-**d**ier, 实验材料为荷兰语), 反应时间显著快于异源组。在其他语言中, 也发现了相同的音位促进效应, 为音位加工的心理现实性提供了证据(see Alario, Perre, Castel, & Ziegler, 2007 for French; Damian & Bowers, 2003 for English)。

另一项支持音位加工的证据来自于颜色图片命名范式(coloured picture naming paradigm, Damian & Dumay, 2007, 2009)。在这一范式中, 向被试呈现有颜色的图片, 要求被试又快又准确地报告出颜色+图片名称。Damian 和 Dumay (2007, 2009)操纵了颜色和图片名称的相关, 使得两者具有相同的首音位, 或是完全无关。研究结果发现, 与无关条件相比, 当颜色和图片名称具有相同的首音位时(**r**ed **r**ope), 命名反应时更短。另外, 这种音位启动效应不局限在首音位上, 当颜色和图片名称的重叠音位位于中间元音或末尾辅音时, 同样

[1] 值得注意的是, 音位(phoneme)与音素(phone)不同。音素是根据语音的自然属性划分出来的最小语音单位。从生理性质来看, 一个发音动作形成一个音素。不同发音动作发出的音, 代表不同的音素。如: 在英语单词 “paper”中, 第一个/p/为送气音[ph], 第二个 p 为非送气清音[p], [ph]和[p]属于同一个音位的不同的音素。音素一般用国际音标记音, 用方括号[]标明。对这些不同的音素, 即使是该语言的母语者甚至可能都觉察不到语音上的差别。在本项目中, 我们关注的是具有抽象特征的音位, 而非音素。

[2] 音位可以分为音段音位(又称音质音位)和超音段音位(又称非音质音位)。超音段音位是具有区别词的语音形式的作用的音高、音强、音长。如汉语中的四个声调是具有区别词的语音形式的音高变化, 称为调位。在本项目中, 我们只关注音段音位, 在这里, 简称为“音位”。

表 1　音位定义以及汉语音位

概念	定义	汉语音位(以汉语拼音字母标示如下)
音位(phoneme)	具体语言中能够区分词义的最小的语音单位，也就是按语音的辨义作用归纳出来的音类。	汉语普通话中有 10 个元音音位(a、o、e、i、u、ü、ai、ei、ui、ue), 22 个辅音音位(b、p、m、f、d、t、n、l、g、k、h、j、q、x、z、c、s、zh、ch、sh、r、ng)。

存在启动效应，从而排除了“词首效应”[3]假说，为音位的心理现实性提供了证据。

2.3　印欧语系言语产生中音位的加工机制

自 20 世纪 90 年代以来，研究者们从多角度对音位的加工机制进行了研究，并取得了一些进展。在内隐启动范式中，研究发现，当反应词仅具有相似的音韵特征(phonological features)时(e.g., /b/和/p/在音韵特征上相似，属于两个不同的音位)，同源与异源条件的反应时不存在差异，排除了“语音特征相似”假说，为荷兰语产生中的音位特异性(而非音韵特征组合)提供了证据。另外，只有当相同音位出现在首音位时，才会产生启动效应；相同音位出现在韵母或是韵尾位置时，不会出现音位启动效应，且启动效应量随着相同音位的数量增加而变大。据此，Meyer (1991)提出，音位的组织是一个从左到右、线性递增的组合方式，从首音到核心元音，最后到尾音部分。

不同的言语产生模型对于音位的位置特定性提出了不同的假设。Dell (1986)模型提出，位于词汇中不同位置的相同音位具有不同的表征，即音位具有位置特定性。如：green 中的/g/与 flag 中的/g/属于两个不同的表征，因为前者为元音前辅音，后者为元音后辅音。Roelofs (1997)提出的 WEAVER 模型则假设，音位以序列的方式从左至右插入音节框架，并不具有位置特定性。“位置特定假设”预期音位启动效应只发生在重复音位在词汇中的位置相匹配时。而“位置非特定假设”则预测无论音位的位置是否匹配，均会产生启动效应。研究结果(Damian & Dumay, 2009)支持了“位置非特定假设”。该研究利用颜色图片命名任务，发现音位在两个词汇中的位置不匹配时(如，**green** frog，相同辅音/g/分别在词首和词尾)，仍表现出相似的音位启动效应，从而支持了音位的位置非特定性(Damian & Dumay, 2009)。

早期考察言语产生时间进程的研究通常采用非自然的语言任务和 ERP 技术相结合的实验方法。例如，利用双重判断任务范式(Dual-Choice Reaction GO/NOGO Paradigm)，van Turennout, Hagoort 和 Brown (1997)以单侧化准备电位(Lateralized Readiness Potential, LRP)为电生理指标，发现语义加工早于语音加工，且每个音位加工大概需要 25 ms。对这种非自然、不发音的按键任务能否探测自然语言产生过程还存在争议，因此研究者试图采用自然发音命名的任务和 ERP 技术相结合的手段对时间进程进行了探讨。Jescheniak, Hahne 和 Schriefers (2003)采用延迟反应的图-词干扰范式发现：语音相关与无关条件诱发的波幅在 400~1000 ms 内有显著差异。Dell’Acqua 等(2010)采用图-词干扰范式，操纵图片名称与干扰词之间的语音相关性，结果发现在刺激呈现后 250~450 ms 的时间窗口内，语音相关与无关条件所诱发的波幅存在显著差异。采用掩蔽启动范式，Blackford, Holcomb, Grainger 和 Kuperberg (2012)发现语音相关与无关条件在 350~550 ms 时间窗口存在显著差异。Indefrey 和 Levelt (2004)采用元分析方法，对 58 个脑电实验进行分析，提出音韵编码所发生的时间进程大致为 275~445 ms。

综上，在印欧语系中，通过多种研究范式，为音位加工的心理现实性提供了佐证，并对音位的加工机制进行了探讨，研究结果证实了音位的特异性、位置非特定性，从左到右、线性递增的组合方式以及时间进程。

2.4　汉语言语产生中音位的心理现实性

较之字母语言，汉语具有独特的音韵特点和正字法结构：a)在汉语中，每个汉字对应着一个音节。与大部分字母语言相比，汉语的音节数量相对较少(大约 400 个左右)；b)在字母语言中，音位信息对应着字母或字母组合(/b/-b, /f/-ph)，即音位

[3] 词首效应，指效应来自于词首字母的重要性，如：语误分析中的“词首效应”，表现为词首音位比其他位置的音位发生语误的频率更高(Dell, 1984; MacKay, 1972)。在词汇识别领域，同样发现，词汇的首字母对词汇的识别比其他位置的字母更重要，如首字母转置对词汇识别的干扰要显著大于词中间或词尾转置(Rayner, White, Johnson, & Liversedge, 2006; White, Johnson, Liversedge, & Rayner, 2008)。

有其正字法上的表征。而汉语中的音位信息没有对应的笔划或笔划组合，即汉语音位没有正字法上的表征; c)古代研究汉语音韵的学者根据汉语音节结构的特点，将音节划分为声母和韵母; d)汉语母语者可以分辨出不同音位带来的差异，如汉语母语者可以觉察出/mi/与/ni/的差异(黄伯荣，廖序东, 2011)。不同语言之间在音韵和正字法特征上的显著差异，可能导致语音加工单元的跨语言差异。而上述的音韵与正字法特征可能导致汉语的语音加工单位为“具有音位表征的音节” (phonemically specified syllables, O’Seaghdha, Chen, & Chen, 2010)。

近年来，研究者采用行为实验对汉语言语产生过程中的语音加工单元进行了研究。与拼音文字的研究结果不同，多数汉语的实验证据支持音节单元说。Chen 等人(Chen, Chen, & Dell, 2002; O’Seaghdha et al., 2010)在内隐启动范式中发现，反应词之间的音节重叠产生了显著的促进效应，而音位重叠对反应时的影响有限。You, Zhang 和 Verdonschot (2012)在掩蔽启动范式中，发现亚音节水平上的重叠不存在启动效应，只有音节水平上的重叠才存在启动效应，强调了音节在汉语产生中的心理现实性。根据印欧语言与汉语不同的研究结果, O’Seaghdha 等人(2010)提出，印欧语言和汉语在音韵特点和正字法规则上的差异，造成语言间位于词汇水平下的第一个语音加工单元(proximate unit, “最邻近语音单元”)存在不同。在印欧语系的字母语言中，最邻近语音单元是音位; 在日语中，是音拍; 而在汉语中，是音节。但需要注意的是, O’Seaghdha 并不否定汉语中音位的作用，指出音位在音节的下一层级中提取、加工。

的确，音节和音位加工不是对立的，不具有排他性。目前已有研究通过实验或计算机模拟的方法为汉语言语产生中的音位加工提供了证据。我们前期的一项研究(Qu, Damian, & Kazanina, 2012, 2013)采用颜色图片命名任务，运用 ERPs 技术对汉语言语产生的音韵编码单元进行探究。实验结果表明，尽管音位重复与非重复两种条件下命名反应时的差异不显著，但是，在刺激呈现后 200~300 ms 的时间窗口内，首音位相同条件比无关条件产生更正的波幅，表现为音位促进效应; 而在随后 300~400 ms 的时间窗口内，音位相同条件比无关条件产生显著的负波，这一负波被解释为由于音位相同条件下同一个音位被重复提取，引起言语内部自我监控机制的高负荷而产生的抑制作用。该研究首次为汉语言语产生中的音位加工提供了电生理证据。另外，该研究对方法论有一定的启示。正如文章所阐述，音位促进效应，被随后发生的自我监控抑制效应抵消，从而造成了该研究和以往研究中并未发现音位在行为学上的效应。换言之，传统的行为学测量手段有很多局限性，而脑电技术可以解决单一反应时方法的局限性(Qu et al., 2013; Qu, Zhang, & Damian, 2016)。Yu 等人(Yu, Mo,& Mo, 2014) 采用内隐启动图片命名范式，利用 ERPs 技术，重复了 Qu 等人的发现。Roelofs (2015)采用计算机模拟技术，构建了包含字母语言、日语、汉语在内的言语产生的计算模型。在该模型中, Roelofs 提出，尽管不同语言间的最邻近语音单元(proximate unit)不尽相同，但是所有语言，包括汉语，均包含音位编码。Roelofs 明确提出 “initial segments were actually prepared in the segment-only condition of the WEAVER++ simulations for Mandarin Chinese” (p.13)。这些重要的发现提示：在汉语言语产生中，除了音节之外，音位可能也是语音加工单元之一。

3 问题提出

综上所述，对“言语产生中的音位加工”这一科学问题，相对于印欧语系的字母语言，对汉语的研究基础还非常薄弱。尚待考察的核心科学问题如下：

第一，在汉语言语产生中，音位加工的心理现实性及其影响因素。目前，支持汉语音位加工心理现实性的证据还相对匮乏且存在争议，我们将结合多种实验范式，根据各范式自身的特点，深入、系统地考察在词汇中处于不同位置的音位的心理现实性。另外，音位表征背后的影响因素还尚未考察。基于汉语音位的特点，我们提出了影响汉语音位加工的潜在因素。我们将着重澄清音位表征是否受到第二语言熟练程度、汉语拼音习得、汉语拼音使用经验的影响。阐明音位加工是否受这些因素影响，将有助于回答音位的心理现实性是否具有普遍性的问题。

第二，汉语言语产生中音位的加工机制。目前这一科学问题还基本处于空白状态。本项目将结合汉语的语音特点及音韵特征，采用自然的发

声命名任务, 利用脑电技术, 从音位特异性、位置特定性、音位组合方式、时间进程四个维度考察汉语音位的加工机制。

4　研究构想

(1)汉语音位的心理现实性还存在争议, 且背后的影响因素还尚待考察。

如前所述, 汉语音位的心理现实性还存在争议, 且背后的影响因素还尚待考察。另外, Qu 等人(2012)和 Yu 等人(2014)的研究只是证明了汉语言语产生中首辅音的启动效应, 尚未考察其他位置的音位(如中间元音、尾辅音)。因此, 还需要多范式、全方位、系统地进行考察。我们将分别采用内隐启动范式—操控反应词之间首音位相关或完全无关、掩蔽启动范式—灵活地操纵不同位置的音位相关性(首音位、中间音位、尾音位), 分析比较音位相关条件下和无关条件下的反应时、脑电波幅、脑电地形图, 进一步为汉语音位加工的心理现实性提供佐证。

语言加工通常受人们所使用的具体语言影响, 那么, 音位在汉语中的作用可能受到第二语言的影响, 尤其当第二语言是以音位为基本语音单位的印欧语言时, 音位在汉语中的作用受第二语言影响的可能性更大。在 Qu 等人(2012)的研究中, 被试为在英国读书、母语为汉语的中国留学生, 这些被试所呈现的音位表征, 很可能受到了第二语言(英语)的影响。在后续 Yu 等人(2014)的研究中, 采用了母语为汉语且英语不熟练的大学生为被试。但鉴于英语为大学生的必修课程, 无法完全排除第二语言(英语)的影响。一些研究提示, 第二语言(英语)的使用经验可以加强被试的音位意识。Verdonschot, Nakayama, Zhang, Tamaoka 和 Schiller (2013)利用掩蔽启动命名任务, 考察熟练汉−英双语者的言语产生加工单元。结果发现, 行为学反应时上的启动效应不仅发生在音节水平上, 还发生亚音节的音位组合水平上。因此, 有必要系统考察第二语言对音位表征的影响。在本项目中, 我们拟操纵颜色名称与图片名称之间的首音节相关或完全无关, 采用事件相关电位技术, 操控第二语言的熟练程度, 通过比较高、低熟练水平汉−英双语者在汉语口语产生中的音位效应, 动态地考察影响音位加工的因素。

另外, 汉语拼音广泛应用于汉字注音、汉字语音教学、计算机打字输入等领域。在汉语拼音字母体系中, 采用拉丁字母对应音位(元音、辅音), 并采用拉丁字母通用的字母表顺序。通过拼音字母可以完满地表达汉语普通话里所有的汉字读音。汉语拼音体系中字母与音位的对应关系可能增加人们以音位为单元进行语音编码。正如文中所述, 已有研究表明汉语拼音对音位意识的获得也有显著影响。该研究发现, 与学习过汉语拼音的被试相比, 没有学过汉语拼音的被试很难完成音位增加/剔除任务(Read, Zhang, Nie, & Ding, 1986)。基于此, 我们提出如下问题: 汉语拼音的习得会影响汉语言语产生中的音位加工吗? 在本项目中, 我们将采用颜色图片命名范式, 考察未受过汉语拼音训练的学龄前儿童和未受过汉语拼音训练的成人的音位效应, 并与接受过拼音训练的人群进行比较, 旨在考察拼音习得对音位表征的影响。另外, 汉语拼音的使用经验也可能影响音位表征。事实上, 汉语拼音的使用经验存在很大的个体差异, 这种个体差异很大程度上是由打字输入法的类型导致的: 拼音输入法使用者的拼音使用频率远远大于五笔字型输入法使用者。在本项目中, 我们将比较五笔输入法和拼音输入法使用者的音位效应, 旨在考察拼音使用经验对汉语音位加工的可能影响。总之, 上述因素的潜在影响还尚未明确。阐明这些因素的影响, 将有助于回答音位的心理现实性是否具有普遍性的问题。

(2)汉语音位的加工机制还基本处于空白状态。

不用语言具有独特的语音特点, 基于汉语语音(尤其是汉语音位)特点, 我们提出关于汉语音位加工机制的四个核心研究问题: 第一, 在汉语言语产生中, 音位是否具有特异性(phoneme-specific)? 汉、英在辅音特点上的不同主要体现在英语中绝大多数辅音为清浊成对对比, 如/p/与/b/, /t/与/d/, /k/与/g/等。而在汉语中, 绝大多数辅音是送气与不送气的成对对比, 且绝大多数为清辅音, 只有少数几个为浊辅音(叶蜚声, 徐通锵, 2010)。例如, 汉语普通话中/b/和/p/两个音位的不同在于送气和不送气这一区别性语音特征。汉、英在元音上的不同主要体现在于音长是否是区分音位的要素, 英语中不同音长可以造成音位的区别, 如长元音/i:/与短元音/i/是两个不同的音位, 在不同的英语单词里可以造成意义的不同。而汉语不区分音位的音长, 因此音长不是区分音位的要素。

如前文所述，在荷兰语中发现，音位具有特异性，即只有当音位相同时才会产生启动效应，语音特征相似的不同音位(如：/b/, /p/)不产生显著效应。目前对汉语的音位特异性还尚未考察。在本项目中，我们拟设置音位相同，或音位在发音或声学相似，或完全不同三种条件，考察不同实验条件的音位启动效应。

第二，音位是否具有位置特定性？位于不同位置的音位(如：奶/nai3/-谈/tan2/，声母/n/或韵尾/n/)具有相同的表征，还是具有位置特定性，即位于不同位置的音位分别独立表征？与一些印欧语言相比，汉语音位位置的规则性更强。比如，俄语、英语允许多至三四个辅音直接结合在一起，根据辅音的数量可将音节分为单辅音音节和辅音连缀型音节；而汉语不允许辅音在音节里直接组合在一起。另外，在英语中，对出现在音节末尾的辅音限制较少，除少数几个辅音之外，有 22 个辅音可以做尾辅音。例如，辅音/g/既可以出现在音节首(“green”)也可以出现在音节尾(“frog”)。而在汉语普通话中，对出现在音节末尾位置上的辅音限制很严。绝大多数辅音只能出现在音节首，不能出现在音节尾。能在音节末尾位置上出现的只有/n/、/ng/两个鼻辅音，而/n/是唯一既可以充当音节首又可以充当音节尾的辅音。

语言间音位位置规则的灵活性可能影响音位的位置特异性，音位位置规则性越强，音位的位置特定性可能越强。该假设预期，与印欧语言相比，在汉语中，音位更可能存在位置特定效应。来自英语的研究结果(Damian & Dumay, 2009)支持了“位置非特定假设”。在本项目中，我们将考察汉语中音位的位置特定性，我们将操纵图片名称之间音位重叠在词汇中所处的位置匹配(比如：均为声母/n/)或不匹配(声母/n/或韵尾/n/)。如果汉语音位具有位置特定性，音位启动效应将只会出现在位置匹配条件中，而不匹配条件将不会产生音位启动效应。

第三，音位的组合方式是线性的，还是非线性的？在一个语言中，音位和音位之间的组合规则，构成了这一语言的音位系统。汉语中音位的组合规则有别于印欧语系。印欧语言大多从字母文字的特性出发，按照元音、辅音对音位进行分析归纳。而汉语普通话音位体系存在两种不同的归纳分析方法：一种以元音、辅音划分，另一种以声母、韵母划分，将音位归纳为声母音位、韵母音位(叶蜚声, 徐通锵, 2010)。

如果是线性组合方式，那么在 CVC 音节(辅音+元音+辅音)中，任意相邻的音位组合的心理意义是等同的，也就是说，CV (声母+韵腹)与 VC (韵腹+韵尾)的心理现实性是一样的。这种线性组合假说从根本上排除了处于音位表征层之上的声母、韵母的表征层，取而代之的是从音位表征层直接到音节层。如果汉语中存在声母、韵母的组合方式，那么 VC (韵腹+韵尾)对语言任务的影响将大于 CV (声母+韵腹)，因为 VC (韵腹+韵尾)组成韵母。本项目将通过比较 CV 与 VC 重叠的启动效应，考察音位的组合方式。

第四，汉语各音位的时间进程是怎样的呢？就音节结构而言，英语音节结构比较复杂，且存在重新音节化现象。与之相比，汉语的音节结构较为简单，常用的无调音节数量仅 400 多个(不区分音调)，有调音节数量为 1300 多个，每个汉字对应一个音节，基本不存在重新音节化现象。基于音位在汉语中的重要性，O’Seaghdha 等(2010)在模型中对比了汉、英语音编码的异同，明确指出汉语词汇的语音编码起始于音节加工，然后进行音位编码，而在英语中，语音编码始于音位加工。如果 O’Seaghdha 模型对跨语言音节和音位发生顺序的推论是正确的，即汉语中音节加工早于音位加工，我们预期音节效应发生的时间窗口将早于音位效应。如果各音位之间呈序列加工，预期首音位效应所发生的时间窗口将早于中间音位效应，中间音位效应又将早于尾音位效应。如果为平行加工，预期各条件下的时间窗口将无差异。

目前对汉语音韵编码的时间进程研究多集中于对音节的考察，而针对音位加工时间进程的考察还非常少。Zhang 和 Zhu (2011)采用语言学任务，结合脑电技术，对各音位以及音位与超音位的相对时间进程进行了探讨。通过双重判断任务范式(Dual-Choice Reaction GO/NOGO Paradigm)，要求被试判断所呈现的图片名称是否包含某个音位或超音位信息，用特定手做按键反应。该范式以 N200 作为神经反应信号，发现在首辅音任务下的 N200 早于中间元音 20~80 ms，而中间元音和声调任务下的 N200 时间进程没有差异。基于此，研究者推论：汉语各音位之间是以递增、序列的方式加工的，而音位与超音位是平行加工的。该研

究无疑为音段加工的相对时间进程提供了重要启示。但是, 该范式存在一些局限性: 第一, 该范式是一个基于语言学知识的按键任务, 该任务不要求被试做出言语表达, 因此该任务能否探测自然语言产生过程存在争议(Jansma, Rodriguez-Fornells, Möller, & Münte, 2004)。第二, 一些研究提示, N200 反映的是反应抑制的神经活动(Jodo & Kayama, 1992; Sasaki & Gemba, 1993), 不直接反映音韵编码的神经活动, 因此只能提供音位加工的相对时间进程, 而不是绝对时间进程。Qu 等(2012)将自然发声的颜色图片命名范式与脑电技术相结合, 研究发现, 音位效应发生在 200~300 ms 的时间窗口, 初步为音位加工的绝对时间进程提供了启示。但该研究仅操纵了首音段, 并没有操纵其他位置的音位信息, 导致无法得知各音段、音段与音节加工的时间进程。在本项目中, 我们将利用 ERP 技术, 考察汉语言语产生中音节和音位效应发生的时间进程, 以及首音位、中间元音音位、末尾辅音音位加工的时间进程。

5　理论建构

目前具有重要影响的两个言语产生理论(Dell, 1986; Levelt et al., 1999), 均假设音韵编码阶段(phonological encoding)负责加工处理词汇的音节结构、韵律特征和音位等语音信息, 并且都认为音位是音韵编码的重要加工单元, 单个音节内部的音韵编码是一个增长式的编码过程, 从音节的首音开始到核心元音, 最后到尾音。以上是建立在英语与荷兰语等印欧语系上的语言产生的模型, 而对于作为非字母语言的汉语来说, 其本身具有独特的语言特点, 使得在考察汉语语音加工单元这一问题上必须结合汉语自身的特点加以研究。本项目立足于汉语语言认知, 基于汉语独特的音韵特点, 采用多种行为学范式和事件相关电位技术, 系统地考察汉语言语产生中音位加工的心理现实性。另外, 现有的言语产生理论模型侧重于言语加工过程本身, 而对说话者自身的语言经验、以及这些语言经验对语言加工机制的潜在影响关注较少。本项目以言语产生者为本, 深入考察说话者的语言习得经验和使用经验(汉语拼音和第二语言), 并探讨这些语言经验对音位加工的影响。

综上所述, 语音加工单元具有跨语言的特异性。不同语言(或方言)在音韵特征上存在着显著差异 (Ladefoged, 2001; Ladefoged & Maddieson, 1996), 从而导致不同语言间语音表征的不同。在以荷兰语或英语为实验材料的内隐启动研究证实, 音韵编码是以音位为加工单元, 以从左到右、线性递增的组合方式进行的 (Meyer, 1990; Roelofs, 1996, 1998)。与印欧语言不同, 音位在日语言语产生中的作用甚微, 而音拍(mora)的作用显著 (Kureta et al., 2006)。值得注意的是, 在汉语中, 研究发现, 音节是言语产生主要的加工单元 (Chen et al., 2002; O'Seaghdha et al., 2010; You et al., 2012)。另外, 初步的研究结果提示, 在汉语言语产生中, 除了音节之外, 音位也是语音加工单元之一(Qu et al., 2012; Yu et al., 2014)。已有研究对音段在汉语言语产生中的时间进程进行了初步探索, 发现汉语各音位之间可能是以递增、序列的方式加工的, 而音位与超音位是平行加工的(Zhang & Zhu, 2011)。

基于上述研究成果和汉语的音韵特征, 我们提出初步的汉语言语产生音韵编码的理论框架: 在汉语言语产生中, 位于词汇水平下的第一个语音加工单元(proximate unit, "最邻近语音单元")为具有音位表征的音节。即在音韵编码阶段, 首先涉及音节的提取。值得注意的是, 在该模型中, 音节是在心理词典中存储的, 而不是在线计算的。在心理词典中存储的音节信息被提取之后, 激活该音节的声母、韵母表征层, 最后激活音位和音调信息。音位的提取是从左到右递进式进行的, 与此同时进行音调加工。整个音韵编码过程伴随着自我监控过程, 即说话者通过实时监控言语产生的整个过程, 对可能发生的语误及时予以纠正。本项目将通过实证研究对这一理论框架的主要假设加以验证、修正。在验证、修正理论框架的过程中, 也将对下一步的实证研究有所启示。

参考文献

黄伯荣, 廖序东. (2011). *现代汉语(上册)*. 北京: 高等教育出版社.

叶蜚声, 徐通锵. (2010). *语言学纲要* (第 4 版). 北京: 北京大学出版社.

张清芳. (2008). 汉语单音节和双音节词汇产生中的音韵编码过程: 内隐启动范式研究. *心理学报*, *40*(3), 253–262.

Alario, F. X., Perre, L., Castel, C., & Ziegler, J. C. (2007).

The role of orthography in speech production revisited. *Cognition, 102*, 464–475.

Baudouin de Courtenay, J. N. (1972). *A Baudouin de Courtenay anthology; the beginnings of structural linguistics*. Bloomington: Indiana University Press.

Blackford, T., Holcomb, P. J., Grainger, J., & Kuperberg, G. R. (2012). A funny thing happened on the way to articulation: N400 attenuation despite behavioral interference in picture naming. *Cognition, 123*(1), 84–99.

Caramazza, A. (1997). How many levels of processing are there in lexical access?. *Cognitive Neuropsychology, 14*, 177–208.

Chen, J. Y., Chen, T. M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language, 46*, 751–781.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper & Row.

Damian, M. F., & Bowers, J. S. (2003). Effects of orthography on speech production in a form-preparation paradigm. *Journal of Memory and Language, 49*, 119–132.

Damian, M. F., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language, 57*, 195–209.

Damian, M. F., & Dumay, N. (2009). Exploring phonological encoding through repeated segments. *Language and Cognitive Processes, 24*, 685–712.

Dell, G. S. (1984). Representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(2), 222–233.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 96*, 283–321.

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language, 27*, 124–142.

Dell'Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., & Grainger, J. (2010). ERP evidence for ultra-fast semantic processing in the picture–word interference paradigm. *Frontiers in Psychology, 1*, 177.

Garrett, M. F. (1980). The limits of accommodation: Arguments for independent processing levels in sentence production. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 263–271). New York: Academic Press.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*, 101–144.

Jansma, B. M., Rodriguez-Fornells, A, Möller, J., & Münte, T. F. (2004). Electrophysiological studies of speech production. In T. Pechmann & C. Habel (Eds.), *Multidisciplinary approaches to language production (Trends in linguistics)* (pp. 361–395). Berlin, Germany: Mouton de Gruyter.

Jescheniak, J. D., Hahne, A., & Schriefers, H. (2003). Information flow in the mental lexicon during speech planning: Evidence from event-related brain potentials. *Cognitive Brain Research, 15*(3), 261–276.

Jodo, E., & Kayama, Y. (1992). Relation of a negative ERP component to response inhibition in a go/no-go task. *Electroencephalography and Clinical Neurophysiology, 82*, 477–482.

Kureta, Y., Fushimi, T., & Tatsumi, I. F. (2006). The functional unit in phonological encoding: Evidence for moraic representation in native Japanese speakers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(5), 1102–1119.

Ladefoged, P. (2001). *Vowels and consonants: An introduction to the sounds of languages.* Oxford: Blackwell.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages.* Oxford: Blackwell Publishers.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *The Behavioral and Brain Sciences, 22*, 1–38.

MacKay, D. G. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology, 3*, 210–227.

Meringer, R., & Mayer, C. (1895). *Versprechen und Verlesen: Eine psychologisch-linguistische Studie.* Stuttgart, Germany: Göschen'sche Verlagshandlung.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language, 29*, 524–545.

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language, 30*, 69–89.

O'Seaghdha, P. G., Chen, J. Y., & Chen, T. M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition, 115*, 282–302.

Qu, Q. Q., Damian, M. F., & Kazanina, N. (2012). Sound-sized segments are significant for Mandarin speakers. *Proceedings of the National Academy of Sciences of the United States of America, 109*, 14265–14270.

Qu, Q. Q., Damian, M. F., & Kazanina, N. (2013). Reply to

O'Seaghdha et al.: Primary phonological planning units in Chinese are phonemically specified. *Proceedings of the National Academy of Sciences of the United States of America, 110*(1), E4.

Qu, Q. Q., Zhang, Q. F., & Damian, M. F. (2016). Tracking the time course of lexical access in orthographic production: An event-related potential study of word frequency effects in written picture naming. *Brain and Language, 159*, 118–126.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107*, 460–499.

Rayner, K., White, S. J., Johnson, R. L., & Liversedge, S. P. (2006). Raeding wrods with jubmled lettres: There is a cost. *Psychological Science, 17*(3), 192–193.

Read, C., Zhang, Y., Nie, H. Y., & Ding, B. Q. (1986). The ability to manipulate speech sounds depends on knowing alphabetic spelling. *Cognition, 24*, 31–44.

Roelofs, A. (1996). Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language, 35*, 854–876.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64*, 249–284.

Roelofs, A. (1998). Rightward incrementality in encoding simple phrasal forms in speech production: Verb-particle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 904–921.

Roelofs, A. (2015). Modeling of phonological encoding in spoken word production: From Germanic languages to Mandarin Chinese and Japanese. *Japanese Psychological Research, 57*, 22–37.

Sasaki, K., & Gemba, H. (1993). Prefrontal cortex in the organization and control of voluntary movement. In T. Ono, L. R. Squire, M. E. Raiche, D. I. Perrett, & M. Fukuda (Eds.), *Brain mechanisms of perception and memory: From neuron to behavior* (pp. 473–496). New York: Oxford University Press.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295–342). Hillsdale, NJ: Erlbaum.

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 109–136). New York: Springer.

Trubetzkoy, N. (1969). *Principles of phonology*. Berkeley: University of California Press.

van Turennout, M., Hagoort, P., & Brown, C. M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(4), 787–806.

Verdonschot, R. G., Nakayama, M., Zhang, Q. F., Tamaoka, K., & Schiller, N. O. (2013). The proximate phonological unit of Chinese-English bilinguals: Proficiency matters. *PLoS One, 8*(4), e61454.

White, S. J., Johnson, R. L., Liversedge, S. P., & Rayner, K. (2008). Eye movements when reading transposed text: The importance of word-beginning letters. *Journal of Experimental Psychology: Human Perception and Performance, 34*(5), 1261–1276.

You, W. P., Zhang, Q. F., & Verdonschot, R. G. (2012). Masked syllable priming effects in word and picture naming in Chinese. *PLoS One, 7*(10), e46595.

Yu, M. X., Mo, C., & Mo, L. (2014). The role of phoneme in Mandarin Chinese production: Evidence from ERPs. *PLoS One, 9*(9), e106486.

Zhang, Q. F., & Zhu, X. B. (2011). The temporal and spatial features of segmental and suprasegmental encoding during implicit picture naming: An event-related potential study. *Neuropsychologia, 49*, 3813–3825.

The functional unit of phonological encoding in Chinese spoken production: Study on phonemes

QU Qingqing; LIU Weilin; LI Xingshan

(*Key Laboratory of Behavioral Science, Institute of psychology, Chinese Academy of Sciences, Beijing* 100101, *China*)

(*Department of Psychology, University of Chinese Academy of Sciences, Beijing* 100049, *China*)

Abstract: Speech production studies have demonstrated cross-linguistic differences in the processing units involved in phonological encoding. It has been widely assumed a critical role of phonemes in spoken production of Indo-European languages. Phonemes are normally conceived of as abstractions of discrete segmental speech sounds which can distinguish the meaning between words. For instance, the word "big"

represents a sequence of three phonemes /b/, /i/, /g/. Currently, investigations on the processing units in Chinese spoken production mainly focused on syllables, whereas only few studies concern the role of phonemes. In the present project, we propose to comprehensively tackle the role of phoneme in Chinese speech production, focusing on its psychological reality, potential factors influencing phoneme-based effects, processing mechanism and temporal properties, using both behavioral and electrophysiological techniques. Specifically, we will investigate: 1) whether phonemic processing has "psychological reality" in Chinese speech production, and whether sensitivity to phonemic representations is artificially induced by exposure to English as a second language, experience of Pinyin, or phoneme-based typing input method? 2) how we process phonemes? Specifically, we are interested in whether phoneme-based effects are phoneme-specific and position-specific, how phonemes associate together to form larger units, and the temporal properties of phonemic processing. The findings of the present project will not only improve our understanding of how Chinese speakers produce words in mind/brain, but also provide insights into the construction and development of theoretical and computational models of Chinese speech production. In addition, the findings will provide basis for cross-language comparisons, and the development of scientific teaching approaches of Chinese phonology.

Key words: spoken production; Chinese; functional units in phonological encoding; phoneme; ERPs